Leadership scenarios in prisoner's dilemma game

S.G. Babajanyan[1], A.V. Melkikh[2], and A.E. Allahverdyan[1]*
[1] *Alikhanyan National Laboratory (Yerevan Physics Institute), Alikhanian Brothers Street 2, Yerevan 375036, Armenia and*
[2] *Ural Federal University, Mira Street 19, Yekaterinburg 620002, Russia*

The prisoner's dilemma game is the most known contribution of game theory into social sciences. Here we describe new implications of this game for transactional and transformative leadership. While the autocratic (Stackelberg's) leadership is inefficient for this game, we discuss a Pareto-optimal scenario, where the leader L commits to react probabilistically to pure strategies of the follower F, which is free to make the first move. Offering F to resolve the dilemma, L is able to get a larger average pay-off. The exploitation can be stabilized via repeated interaction of L and F, and turns to be more stable than the egalitarian regime, where the pay-offs of L and F are equal. The total (summary) pay-off of the exploiting regime is never larger than in the egalitarian case. We discuss applications of this solution to a soft method of fighting corruption and to modeling the Machiavellian leadership. Whenever the defection benefit is large, the optimal strategies of F are mixed, while the summary pay-off is maximal. One mechanism for sustaining this solution is that L recognizes intentions of F.

Keywords: leadership; followership; prisoner's dilemma; hierarchic games; mixed strategies

I. INTRODUCTION

The phenomenon of leadership is studied in various disciplines (biology, psychology, sociology, management science) and led to a big literature; see [1–9] for reviews. Still there is a common opinion that the research on leadership lacks integrity [9, 10]. One reason for this is a shortage of transparent mathematical models.

Existing models divide into two groups. Agent-based models focus on entities that can modify their interaction due to leaders [11–16]. In particular, the activity of agents can be mediated and regulated by leaders [11, 12, 16]. Such models can mimic leadership scenarios known in society [16], but the leader-follower interaction assumed by them is frequently oversimplistic; e.g. it lacks notions of fairness, efficiency *etc.*

Within game-theoretical approaches—both in economic [5, 17, 18] and evolutionary [6–8] set-ups—the attention of leadership researchers is focused on Stackelberg's solution in coordination games [5, 6, 8, 17, 18]. This provides the most traditional understanding of leadership, where the leader L makes the first move and thereby imposes a solution of the game on the follower F, which reacts (best-responses) to actions of the leader [30–32]. Stackelberg's solution describes certain leadership scenarios in human communities or in animal groups [6, 8]. But the leadership phenomenon is richer than follower(s) reacting on leader's actions.

Leaders are also studies within evolutionary game-theoretic models on graphs, in particular on complex networks; see [33–35] for reviews. Here leaders are associated with strongly connected nodes (i.e. hubs) of the underlying network. Hubs can influence the behavior of other nodes. But relating hubs with leaders is incomplete [36], since static networks do not contain information on the dynamics of influences, e.g. on information transfer from one node to another [36].

We study the prisoner's dilemma game that has many realizations in economics, politics and social relations [19–27]; see section II A for a reminder. Here the jointly effective (cooperative) actions of both player are unstable with respect to a unilaterial change (defection), which leads both players to inefficient (jointly defecting), low pay-off state. Yet this state is robust, since it is a Nash equilibrium of the game. It is also obtained via Stackelberg's solution, or via the concept of dominant strategies. Resolving the prisoner's dilemma amounts to finding mechanisms that can lead to avoiding this inefficient state, thereby going beyond simple rationality ideas related to the Nash equilibrium or dominant strategies [20–26].

Our aim here is to work out a class of leadership scenarios that emerge from within the prisoner's dilemma game. The previous work on leadership in social dilemma games studied an exogenous leader that initiates cooperation between several (at least 3) followers engaged in a prisoner's dilemma-type multi-round game [28]. In contrast, the present work will focus on the two-player prisoner's dilemma game, where one of players shows features of a transactional and transformational leader. Our interest will be in psychological and game-theoretic mechanisms that enforce such leadership already in a one-shot (single-round) situation. Here is a brief description of our model and its main results.

---

* armen.allahverdyan@gmail.com

(i) We describe a leadership scenario, where L commits to react probabilistically to actions of F, which makes the first move; see section II B. The reaction probabilities of L are determined from the conditional maximization of the average pay-off of L, which is conditioned by the fact that the average pay-off of F is larger than its guaranteed value. Under certain conditions this leadership scenario leads to Pareto-optimal solutions; see section IV. The scenario is similar to transactional leadership [1], since the response of L amounts to encouragement—if F cooperates, then F gets a higher average pay-off than the guaranteed value—and discouragement, where the pay-off of non-cooperating F is reduced to its guaranteed value. Also, the situation has basic features of the transformative leadership [1]: *(1.)* F is free to make the first move and accept the game rules [2]. *(2.)* L is able to take F out of the non-cooperative behavior, i.e. out of the dominant strategy of F.

(ii) The model has two regimes: egalitarian *versus* exploiting leadership. In the first regime L helps F to resolve the prisoner's dilemma, but gets the same *average* pay-off as F. However, *generically* L exploits F, since the average pay-off of L is larger. Importantly, this regime is more stable than the egalitarian one. Note that the leadership scenario is to be stabilized with respect to L breaking his commitment (i.e. deceiving). One—but not the only—stability mechanism comes from a repeated-game implementation of the scenario, where different rounds of the game are independent and identically distributed; see section II C. There the stability comes from a back-reaction of F on L (i.e. F punishes L), when L breaks the commitment. F is well-motivated to punish L, since F looses less out of this punishment; this mechanism is absent in multi-game implementations of the usual prisoner's dilemma. Thus the model provides a mechanism for transition from egalitarian leaders to exploiting ones, a widely discussed problem in the leadership research [5–8].

(iii) When the pair $\mathrm{L}+\mathrm{F}$ is viable as a whole, i.e. when their summary pay-off is large? The summary pay-off can be related to group fitness [24] or, alternatively, to the overall resource extracted from the environment. Hence this question concerns the mission of leaders that are supposed to organize efficient, competitive groups [9]. The model shows that with respect to the summary pay-off, the exploiting regime is never better than a certain egalitarian one (which can be more complex in its implementation); see sections II D and IV.

(iv) The exploiting regime assumes a follower F that agrees to get less than L, and is supposed to cooperate with L, while L will defect F with a certain probability. Though F does have a freedom of defecting, he will be discouraged (punished) by defection in response. A closer look to the situation shows that the exploiting leader has features of manipulative (Machiavellian) personality, as described in experimental literature [40–42] [1]. Hence it can serve as a game-theoretic formalization of the Machiavellian leadership [43, 44]; see section V. How F can evolve given an exploiting L? One possibility is that a greedy (pay-off maximizing) L will lead to an apathetic F, whose actions have nearly the same average pay-off. If the defection benefit is high (i.e. $T > 2R$ in (1)), L can allow F to employ a mixed strategy, i.e. to defect with a definite probability without necessarily punishing him; see section IV C. This solution is Pareto-optimal for $T > 2R$, but it is also more complex and difficult to manage, because it requires that L recognizes intentions of F; see section IV. Once for $T > 2R$ the complexity is managed and the scenario is implemented, L does not have to exploit F generically, and also the overall pay-off of $\mathrm{L}+\mathrm{F}$ is equal to its maximal possible value.

We organized this paper as follows. Section II A reminds the prisoner's dilemma game and sets notations. Section II B presents the main leadership scenario for resolving the prisoner's dilemma, discusses its stability, and the summary pay-off. Section II E provides a realistic example of this scenario based on a soft method of fighting corruption in developing countries. Section III discusses relation with literature. Section IV deduces the leadership scenario from a more general set-up, where F can act probabilistically (i.e. within a mixed strategy). Section IV also studies the prisoner's dilemma game in the regime, where the defection benefit is high (i.e. $T > 2R$). Section V summarizes our results in the context of open questions in the leadership research. A minimalist reader interested only in basic implication of our results for leadership theories can read sections II A, II B, and V.

---

[1] It is also well-known that animal leaders manipulate their followers in order to gain more [5].

II. LEADERSHIP SCENARIOS FOR THE PRISONER'S DILEMMA GAME

A. The game and the dilemma: a remainder

There are two players, L and F. Each one can apply two actions: c (cooperate) and d (defect). Pay-offs are determined by the following matrix

L/F	c	d
c	R, R	S, T
d	T, S	P, P

, (1)

where e.g. actions c and d by (resp.) L and F result to pay-offs S and T to (resp.) L and F. The prisoner's dilemma game is specified by the following relation between the pay-offs:

$$T > R > P > S. \tag{2}$$

Since all pay-offs can be simultaneously multiplied by a positive number and added an arbitrary number, we choose without loss of generality:

$$R = 1, \qquad S = 0, \qquad \text{hence} \quad 1 > P > 0. \tag{3}$$

Eqs. (1, 2) show that for both players d is a dominant strategy, i.e. d yields a higher pay-off than c, no matter what the opponent does. It follows that no player can gain by unilaterally changing his strategy only if both defect; i.e. (d, d) is the only Nash equilibrium of game (1). Would both acted c, they would both get R, which is larger than P in the Nash equilibrium. But playing c is vulnerable, since the opponent can change to defecting, gain out of this, and leave the cooperator with the minimal pay-off $S = 0$. This makes the famous prisoner's dilemma [20–23].

Eq. (1) defines a symmetric game: both players are equivalent on the level of pay-offs and actions. One can try to break this equivalence on the level of behavior, i.e. making L a leader, and F his follower. Since the game is symmetric, the asymmetry between L and F refers to their attributes, i.e. it is not imposed by pay-offs (1) of the game.

One leadership scenario is notoriously unsuccessful in solving the dilemma. It is based on the concept of Stackelberg's solution [30–32], which is a game-theoretic manifestation of coercive (or autocratic) leadership [4]. Within this solution, L makes the first move, knowing that F will respond by his best reponse to every move of L. Since the best response of F is always d, both players will end up in (d, d). Stackelberg's solution is also inefficient for the sequentially played prisoner's dilemma game, where it is also known as the backward induction [21].

B. Solution via conditional probabilistic commitment

We propose to solve the dilemma by keeping the idea of leadership, but now the leader L does not impose a solution on the follower F. In contrast, F is given freedom to decide for himself, while L provides F with probabilities of his reactions to various actions of F (i.e. L commits to a probabilistic reaction [45]): if F chooses d, then L will take d in response. If F will act c (cooperation), then L will cooperate with probability $1 > P + \delta > 0$:

$$\Pr[\mathrm{L} \mapsto d \,|\, \mathrm{F} \mapsto d\,] = 1, \quad \Pr[\mathrm{L} \mapsto c \,|\, \mathrm{F} \mapsto c\,] = P + \delta, \quad \delta > 0, \tag{4}$$

where $\mathrm{L} \mapsto d$ means L acting d, $\Pr[...\,|\,...]$ means conditional probability, and δ is a parameter to be discussed below. If F acts c, then L and F get (resp.) the following *average* pay-offs:

$$\mathfrak{L}_c = (P + \delta) \times 1 + (1 - P - \delta) \times T, \tag{5}$$
$$\mathfrak{F}_c = P + \delta. \tag{6}$$

If F acts d, then both get

$$\mathfrak{L}_d = \mathfrak{F}_d = P. \tag{7}$$

For F it is beneficial to cooperate (act c), since in average he gets more than for defecting: $\mathfrak{F}_c > \mathfrak{F}_d$ due to $\delta > 0$. This is why F chooses to act c. Eq. (5) shows that $\mathfrak{L}_c$ increases with decreasing δ. Hence a greedy L would prefer to make δ smaller, but for $\delta \to 0$ the situation becomes meaningless, since c and d are of equal value for F.

According to (5, 6) there is a clear asymmetry between L and F: L gets (in average) larger than R, $\mathfrak{L}_c > R = 1$, while F gets smaller than R, $\mathfrak{F}_c < R = 1$, but still larger than for the defecting strategy ($\mathfrak{F}_d = P$). Below we shall see that this asymmetry can make the situation more stable. Note that for following these rules of the game, L has to known all pay-off of the game, while F may—but need not—know pay-offs of L. For simplicity and definiteness we shall assume that F knows the pay-offs of L.

Section IV shows that (4–7) come out from an optimization principle, where L tries to maximize his *conditional average* pay-off; i.e. the average pay-off given that F acts c and gets an average pay-off larger than P. Also, provided $T < 2R = 2$, solution (4–7) is Pareto-optimal, i.e. it is impossible to increase the average pay-offs of *both* L and F by deviating from it; see section IV. We emphasize that although we are looking at a single-shot game, the concept of average pay-off does apply; see e.g. [21].

What will prevent L from deceiving, i.e. announcing probabilistic response via (4), but acting d (i.e. his best response) with probability one? One answer to this question relates to the repeated game implementation of (5–7) and is discussed below. Another possible answer is that the leader L will have some prestige losses if his initial promise is not followed. Then the solution is stable if such losses are larger than $T - \mathfrak{L}_c$. Hence the asymmetry between L and F can make the situation more stable, because $T - \mathfrak{L}_c$ can be sufficiently small. Note that the same story on prestige can be told for the ordinary prisoner's dilemma; cf. section II A. But there it amounts to a stronger assumption, since each player should be assumed to have prestige (or reputation) losses larger than $T - R$, and the summary prestige loss should be larger than $2(T - R)$. Now $2(T - R)$ is larger (and can be much larger) than $T - \mathfrak{L}_c$. Moreover, introducing prestige losses for the ordinary prisoner's dilemma amounts to a trivial change of pay-offs in (1) that amounts to making the outcome (c, c) the Nash equilibrium of the game.

Notions of reputation and prestige became recently popular for explaining aspects of human cooperative behavior [35]. The relevance of these notions is supported experimentally, e.g. it is known that people whose behavior is observed tend to cheat less [35]. Ref. [46] provides a fuller account of various prestige (reputation) factors; see also [49, 52] for examples and [35, 50, 51] for reviews.

C. Implementation via repeated games

1. *Stability against deception*

So far we focused on the single-shot game described by (1–7). Recall that according to the law of large numbers, the averages in (5, 6) can be realized as actual pay-offs in sufficiently many identical, independent rounds of the game. We assume that the action of L—in particular, if this action is a deception, i.e. acting d, despite of the fact that the probabilistic mechanism governing (4) generated c—becomes known to F before the next round of the game. More realistically, F will have to learn about deceptions of L via looking at a sufficiently long statistics of actions of L.

Importantly, we assume that L and F will play $N = N_1 + N_2$ times, where $N \gg 1$ is a sufficiently large number that is not known to L and to F [2].

---

[2] In particular, even immediately before playing the last round of the game, L does not know that it will be the last play. This assumption is not be met in many real-life examples of repeated games. However, it is inavoidable, since without this assumption the situation is vulnerable to the backward induction, which essentially trivializes the situation for the prisoner's dilemma game [21]. Indeed, knowing that a given round is the last one, L (which moves the last one) will certainly act his best-response against any move by F. Now F will certainly know this, so he will know what to play in the last round *etc.* So without the above assumption (and if other assumptions are not made), we shall always be confined by the situation, where both L and F act d, i.e. the prisoner's dilemma holds.

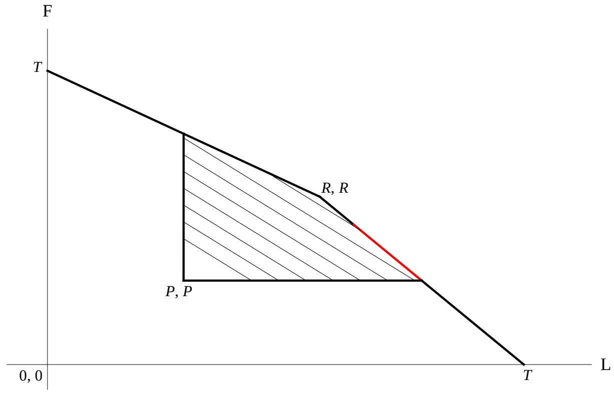


FIG. 1: Pay-off diagram of prisoner's dilemma game under conditional probabilistic commitment solution; see section II B. Pay-offs of L (x-axes) and F (y-axes) for prisoner's dilemma parameters $T = \frac{7}{4}, R = 1, P = \frac{1}{2}$ in (1, 2). At these particular values of parameters, $\mathfrak{L}_c$ and $\mathfrak{F}_c$ are given by the red line; the end-points of this line are $\delta = 0$ and $\delta = \frac{T-P}{2(T-1)} - P$. The later value of δ saturates (8) at $\chi = 1$. It is seen that the summary pay-off $\mathfrak{L}_c + \mathfrak{F}_c$ is smaller than $2R = 2$.
The stroked region represent individually rational and feasible pay-offs that appear in the folk theorem for sequentially played (multi-round) game, where each round is governed by (1). It is seen that the red line is a part of the Pareto front.

Due to $N \gg 1$, the pay-off accumulated by L (resp. by F) is approximately equal to average pay-offs times N. We assume that after each deception, F opposes L (punishes L) by acting χ times d. Let us assume that there are N_1 cases, where F had to respond by c, but acted d (i.e. deceived his commitment). If L deceived N_1 times, he will be opposed $N_2 = \chi N_1$ times by F. This is compared with a case, where L never deceives, and the game goes on N times, i.e. the gain of L (F) is $N\mathfrak{L}_c$ ($N\mathfrak{F}_c$); cf. (5, 6). Now if

$$N_1 T + N_2 P < (N_1 + N_2)\mathfrak{L}_c \qquad \text{or} \qquad T + P\chi < (1+\chi)\mathfrak{L}_c, \tag{8}$$

then deceiving is not beneficial. For a sufficiently large χ, (8) always holds due to $\mathfrak{L}_c > R = 1 > P$ in (5). If F never opposes, $\chi = 0$, then (8) reduces to $\mathfrak{L}_c > T$ which can never be satisfied, since the average pay-off cannot be larger than the maximal one. Note that (8) can be also satisfied for $\chi < 1$, which can be interperted by saying that F opposes with probability χ.

It is seen from (8) that, if the influence of F on L is limited, i.e. when χ cannot be sufficiently large, there are situations where the fair strategy (with $\delta = 1 - P$) is not stable, while the unfair strategy with $\delta < 1 - P$ is stable. In such situations, F may prefer to accept the unfair rules with a partial defection of L, because this makes the situation more stable with respect to deception.

We emphasize an important aspect of above opposing (punishing) by $\mathfrak{F}$: it is well-motivated specifically due to the asymmetry between L and F: whenever F opposes L, F knows that he will loose $\mathfrak{F}_c - P$, which is smaller than what L will loose, i.e. $\mathfrak{L}_c - P$ [3]. Note that the punishment for the usual (symmetric) multi-round prisoner's dilemma game does lack such a motivation. Here the punishing player looses as much as the punished one.

2. *Stability with respect to sequential defection*

For the sequentially played prisoner's dilemma game it is frequently emphasized that the situation is non-trivial only under an additional assumption $2R = 2 > T$ in (2) [22]. Otherwise, there is an additional cooperation scenario that shows itself in an even number of game rounds: the two players can defect each other consequently, i.e. to act (c, d) in the first tour and (d, c) in the second tour. This leads to sharing the pay-offs. In this way each one gets $\frac{T}{2}$ per round, and for $2R = 2 < T$ this is larger than R obtained for (c, c).

For our situation the condition $1 > \frac{T}{2}$ is not critical in the sense that even for $1 < \frac{T}{2}$ the average pay-off (5) of L need not be smaller than $\frac{T}{2}$. Hence L is not interested to change the situation and the probabilistic solution is stable

[3] This mechanism of motivating punishment was described in Ref. [19] (section 5.10), but to our knowledge it was not applied to prisoner's dilemma.

with respect to playing $(c,d)+(d,c)$. Indeed, we note that from (5) $\mathfrak{L}_c > \frac{T}{2}$ leads to an upper bound on δ:

$$\delta < \frac{P + T(\frac{1}{2} - P)}{T-1}. \tag{9}$$

For $T > 2$ the RHS of (9) is smaller than $1-P$; cf. (4). Hence $T > 2$ and (9) impose a non-trivial upper bound on δ that can be satisfied for a sufficiently small δ.

D. Summary pay-off

One can look at this situation from a global viewpoint, i.e. comparing the two leadership strategies—fair vs. exploiting—with respect to the summary pay-off. Hence we compare $\mathfrak{L}_c + \mathfrak{F}_c$ from (5, 6) with $2R = 2$, which is the summary pay-off for the fair situation. One motivation for doing this comes from the group selection ideas (in biology, ecology and economics) [24], where the summary pay-off may play the role of fitness if the dyadic group (L + F) is looked up as a whole and compared with other groups.

We get from (5, 6)

$$\mathfrak{L}_c + \mathfrak{F}_c - 2 = (T-2)(1-P-\delta), \tag{10}$$

i.e. for $T < 2 = 2R$, the exploiting leadership has a smaller summary pay-off than for the cooperative strategy (c,c). Now for $T > 2$, we get $\mathfrak{L}_c + \mathfrak{F}_c > 2$, but for $T > 2$ there is (for even number of games) another fair and symmetric strategy, *viz.* $(c,d)+(d,c)$ that achieves the summary pay-off equal to T, as we saw above. Then we have

$$\mathfrak{L}_c + \mathfrak{F}_c - T = -(T-2)(P+\delta), \tag{11}$$

which is again negative for $T > 2$. We conclude from (11) that, at any rate, there is always a fair solution that outperforms the considered probabilistic solution in terms of the summary pay-off.

E. Application: fighting corruption softly

We describe here a new applications of the prisoner's dilemma game (1, 2), where the solution (5, 6) is realistic. Let L be a potentially corrupt official (bribee) with actions c and d referring to, respectively, not taking and taking bribes. F is the government that decided to eradicate corruption via raising salaries of officials by paying them wage bonus. Such anti-corruption policies were implemented in certain developing countries, e.g. in Syria during early 1980's [56] and in Armenia more recently [55]. There is a clear empiric evidence across various developing countries showing that low wages can indeed promote corruption [57]. Hence F have to actions: to keep the wage bonus (c) or to skip it (d). From the viewpoint of L in (1): P is the bribe money, $R = 1$ is the bonus and $T = R + P < 2R$. The pay-offs of F in (1) are interpreted in a similar way, but they are more convoluted, since they involve both money and consequences of corruption. Hence the game need not be symmetric with respect to L and F, but for clarity we stick to the symmetric situation (1), also because consequences of corruption are difficult to estimate quantitatively.

Now (5, 6) refers to the situation, where the government F does pay the bonus, but L will still take bribes with a certain probability. Note that (10) applies due to $T = R + P < 2R$, showing that the summary pay-off is smaller than $2R = 2$, i.e. this method is not efficient from a global viewpoint. To our understanding the solution described by (4, 5, 6) was qualitatively realized in Armenia: extra wages to officials were never withdrawn (i.e. F always acts c), and the corruption was controlled but not eradicated; e.g. the country's transparency index is still low and did not improve [58].

Note that when officials overdo with bribing (i.e. they raise the probability of defection), the government prosecutes them legally. At least, this was the main scenario of government actions in Armenia. Hence within this realization of the prisoner's dilemma game and its probabilistic commitment solution [cf. section II B], the punishment of the deceiving leader involves—unlike the repeated implementation of the prisoner's dilemma game discussed in section II C—actions that go essentially out of the prisoner's dilemma game itself.

III. RELATIONS WITH LITERATURE

A. General remarks

The solution (5, 6) does have several important predecessors. To our knowledge, the first example of conditional probabilistic commitment was mentioned by Schelling [45], who discussed it in the context of dominated strategies of

L; see [46] a recent broad discussion of various aspects of commitment. Yet another related approach was proposed by Brams [47], who studied the prisoner's dilemma in the context of the Newcomb's paradox; see also [48] in this context. The probabilistic prediction scheme developed in [47, 48] can be regarded as a particular case of the conditional probabilistic commitment studied below. The solution (5, 6) also shares certain features of adaptive strategies found in [26, 27, 29]; see also [50] for discussion. However, these solutions are completely symmetric with respect to players. In particular, both of them act simultaneously. For those solutions $T < 2R$ is a critical condition, in contrast to our situation; see section II C 2.

The fair (deterministic response) solution (4) coincides with the one studied by Howard [37], within the so called inverse Stackelberg's solution; see [38, 39] for recent reviews. Note that Howard also proposed the so called second-order metagame, whose application to the prisoner's dilemma game lead to non-trivial predictions and was considered to be a solution of the prisoner's dilemma [20]. However, the general approach of the second-order metagames is convoluted and involves hidden assumptions; see Appendix B for a short discussion.

B. Relations with the ultimatum game

Solution (4–6) has certain similarities with the ultimatum game; see e.g. [62, 63] for in-depth description of this game. We shall discuss these similarities with two purposes in mind: first the ultimatum game can provide some information on the partially constrained parameter δ in (4–6). Second, the ultimatum game itself gives a clear example, where people do not employ their best-response action.

Within the ultimatum game, the first player I proposes to the second player II to divide a certain amount of money $\mathfrak{M}$ according to two different strategies. The fair strategy f gives $\mathfrak{M}/2$ to each player. The second strategy u gives $0.01\,\mathfrak{M}$ to II, while I gets $0.99\,\mathfrak{M}$. Now II can just accept the proposal by I (action a) or reject it (action r). In the latter case, both I and II get nothing. Since II reacts on actions of I, the strategies of II are aa (always accept), rr (always reject), ar (react a on f and r on u), and ra. Using (A4) from Appendix A, one can show that the pairs of strategies (f, ar) and (u, aa) are Nash equilibria, but only (u, aa) is the subgame-perfect Nash equilibrium, since (ever) playing r is against the interests of II: better a puny pay-off than nothing [62, 63]. The concept of sub-game perfectness postulates that if actually playing r is not rational, then also treatening to play r is not rational. This postulate is (expectedly) invalidated in experiments with human subjects playing the ultimatum game, which routinely act r in response to u [62, 63]. The pair (u, aa) is seen experimentally only if the fraction 0.99 is replaced by number closer to (but still smaller than) 0.5; e.g. 0.3 [62, 63].

The analogy of this situation with (4–6) is that there as well L proposes to F a deal that is asymmetic (i.e. L generically benefits more than L), but does neverthelss provide to F an average pay-off $P + \delta$ larger than its guaranteed value P. Thus by analogy with experiments on the ultimatum game, we expect that human followers F will not accept small values of δ, instead choosing to act d. Note that d is the dominant strategy for F, which is not the case with action r for the second player in the ultimatum game. Moreover, even if F agrees with the deal, he may still be deceived by L, who acts the last one. These arguments strengthen the point against small values of δ. Another difference is that in the ultimatum case the summary pay-off is always fixed and equals its maximal value $\mathfrak{M}$, in contrast to (4–6), where the summary pay-off is generically smaller than $2R$; cf. (10).

C. Relations with folk theorems for repeated games

General features of our implementation of the probabilistic commitment solution via repeated games (see section II C) are governed by the folk theorem [21, 60, 61]. Hence we shall review it and specify its relation with our results. We shall focus on the implementation of repeated games that is similar to what we discussed in section II C. Here the average pay-offs are just arithmetic averages of single-round pay-offs, which—due to the law of large numbers and for sufficiently long series of rounds— coincides with probabilistic averages seen within the single-shot realization of the game [4]. Then the main (folk) theorem of the set-up starts by looking at the convex sum of individual pay-offs; e.g.

[4] Note that there is another mechanism for implementing repeated games. It goes by introducing the discount factor, which is a finite probability by which every round of the repeated game takes place. This factor allows to get finite average pay-offs for a long series of rounds [21, 60, 61]. We shall not employ this implementation, because the discount factor (being an additional parameter) makes pay-offs essentially different from their single-shot analogues.

for the prisoner's dilemma game (1) these are vectors of the form

$$\alpha_1 \times (R, R) + \alpha_2 \times (S, T) + \alpha_3 \times (T, S) + \alpha_4 \times (P, P), \tag{12}$$

$$\sum\nolimits_{k=1}^{4} \alpha_k = 1, \qquad \alpha_k \geq 0, \qquad k = 1, ..., 4, \tag{13}$$

where α_1 is the probability by which the action (R, R) is taken *etc.* The first (second) component $\alpha_1 R + \alpha_2 S + \alpha_3 T + \alpha_4 P$ $(\alpha_1 R + \alpha_2 T + \alpha_3 S + \alpha_4 P)$ of vector (12) refers to the pay-off by L (F). The convex domain defined via (12) is to be bound by restricting both components of (12) to values larger or equal than P; see the stroked domain in Fig. 1 (for $2R = 2 > T$) and in Fig. 3 (for $2R = 2 < T$). Now the folk theorem states that every point within the (stroked) domain can be realized via a certain joint strategy L and F in the repeated game as a a sub-game perfect Nash equilibrium [60]. This strategy (which is generally not unique) amounts to to L and F acting (c, c) for the first n_1 rounds, then acting (c, d) for subsequent n_2 rounds, acting (d, c) for further n_3 rounds, and (d, d) for n_4 rounds. Then the cycle is repeated, i.e. (c, c) for n_1 rounds *etc.* Here $N = M(n_1 + n_2 + n_3 + n_4) \gg 1$ $(M \gg 1)$ is the overall number of rounds, and the relation with the average pay-off vector (12) is obvious: $\alpha_k = \frac{n_k}{n_1+n_2+n_3+n_4}$. If one of players deviates from this agreement, another one punishes him by acting d for the rest of the multi-round game. This is the Nash equilibrium, since none of players will benefit by deviating from this strategy unilaterially. It is sub-game perfect, since it still decribes a Nash equilibrium if the game starts from ith round ($i > 1$, but $i \ll N$) instead of the first round [60].

Note that the above construction was realized as deterministic game for each round, and we also did not specify the order by which L and F act within each round. In contrast, within the implementation in section II C, L acted probabilistically after F. These differences are not essential, as far as the above folk theorem and its implications are concerned. This is seen from the fact that (12) describes an averaged pay-off.

Hence according to the folk theorem, all points within the stroked domain in Figs. 1 and 3 have equal status. The folk theorem is clearly silent about additional mechanisms that are needed for selecting any specific point or sub-domain there. This is a weak point of the folk theorem [60]. This point gets worst by noting that the stroked domain is not even restricted to Pareto efficient pay-offs for L and F. In contrast, our solution—which is also realized by sub-game perfect Nash equilibrium strategies, as we saw in section II C—restricts within the stroked domain a sub-set of Pareto efficient pay-offs; see the red line in Fig. 1 and the blue line in Fig. 1. The upper end point of this line is determined by parameters δ and χ; see section II C.

IV. GENERAL CASE

A. Definition

So far we postulated (4) and studied its consequences. Now relations in (4) will be deduced from a general setting, where F can employ mixed strategies in his first move. We shall then see that (4) emerges as a Pareto-optimal solution for $T < 2R$; cf. (1).

Now F makes the first move (act) by choosing between two mixed strategies. Within the first strategy, F acts c with probability x and is responded by L with conditional probabilities p_1 and p_2:

$$p_1 = \Pr[\mathrm{L} \mapsto c \,|\, \mathrm{F} \mapsto c\,], \quad p_2 = \Pr[\mathrm{L} \mapsto c \,|\, \mathrm{F} \mapsto d\,], \quad x = \Pr[\,\mathrm{F} \mapsto c\,]. \tag{14}$$

Within the first strategy, F acts c with probability y and is responded by L with conditional probabilities q_1 and q_2:

$$q_1 = \overline{\Pr}[\mathrm{L} \mapsto c \,|\, \mathrm{F} \mapsto c\,], \quad q_2 = \overline{\Pr}[\mathrm{L} \mapsto c \,|\, \mathrm{F} \mapsto d\,], \quad y = \overline{\Pr}[\,\mathrm{F} \mapsto c\,]. \tag{15}$$

This set-up is more complex than the previous case, where $x = 1$ and $y = 0$. For now L has to distinguish from which strategy the defection of F came. E.g. this can be done via controlling the probabilistic mechanism of actions of F. Another possibility is that L is aware of intentions of F when acting d, i.e. L has side (e.g. emotional) reasons to believe whether d was really acted within the mixed strategy [5].

But this more complex set-up has its benefits, because—as seen below in section IV C—it can improve the pay-off for both L and F, for $T > 2R$. If L is not ready to bear this complexity, he will likely oppose any non-deterministic strategy of F by treating all defections in the same way. Such a leader will stick to solution (4).

[5] For recent discussion on intentions in game theory see [64, 65].

B. Derivation

L determines p_1, p_2, q_1, q_2, and y in (14) from maximizing his average pay-off conditioned upon the fact that F gets more within (14) than within (15) [cf. (1, 2)]:

$$\mathfrak{L} = \max{}_{0\le p_1,p_2,q_1,a_2,y\le 1}[\; p_1x\times 1 + p_2(1-x)\times 0 + (1-p_1)x\times T + (1-p_2)(1-x)\times P, \tag{16}$$
$$p_1x\times 1 + p_2(1-x)\times T + (1-p_1)x\times 0 + (1-p_2)(1-x)\times P > \tag{17}$$
$$q_1y\times 1 + q_2(1-y)\times T + (1-q_1)y\times 0 + (1-q_2)(1-y)\times P], \tag{18}$$

where the maximization over x will be caried out later on (for the time being it is interesting to leave it as a free parameter). Note that the average pay-off of L, given by the right-hand-side of (16), does not depend on q_1, q_2 and y. Hence the maximization of (16) is achieved whenever the restriction (18) is possibly weak, i.e. $q_1y + q_2(1-y)\times T + (1-q_2)(1-y)P$ is to be minimized. But the latter quantity—which is the average pay-off of F within (15)—cannot be smaller than the guaranteed pay-off P of F. Otherwise, (15) is meaningless for F. Hence we should have $q_1y + q_2(1-y)\times T + (1-q_2)(1-y)P = P$, which is achieved for

$$q_1 = 0, \quad q_2 = 0, \quad y = 0, \tag{19}$$

i.e. the optimal—from the viewpoint of L—choice of parameters in (15) amounts to F defecting with probability 1, achieving the guaranteed pay-off, and then defected by L in response.

Thus (16, 18) amount to

$$\mathfrak{L} = \max{}_{0\le p_1,p_2\le 1}[\; p_1x + (1-p_1)xT + (1-p_2)(1-x)P, \tag{20}$$
$$p_1x + p_2(1-x)T + (1-p_2)(1-x)P > P\;]. \tag{21}$$

It is now clear that assumptions (14, 15) on just two different strategies is not essential: any number of them will lead to (20, 21) via the same mechanism.

Constraint (21) can be written as

$$\Delta \equiv p_1x + p_2(1-x)(T-P) - xP > 0, \tag{22}$$

where Δ is bounded from above due to $1\ge p_1$ and $1\ge p_2$:

$$\Delta \le (1-P)x + (1-x)(T-P). \tag{23}$$

It appears that the straightforward maximization in (20, 21) leads to $\Delta = 0$. This is meaningless, since then F will not be interested to do anything but defection. Hence we shall take $\Delta > 0$ as a parameter of the solution, express p_1 through Δ and compute (20) by maximizing over a single parameter p_2:

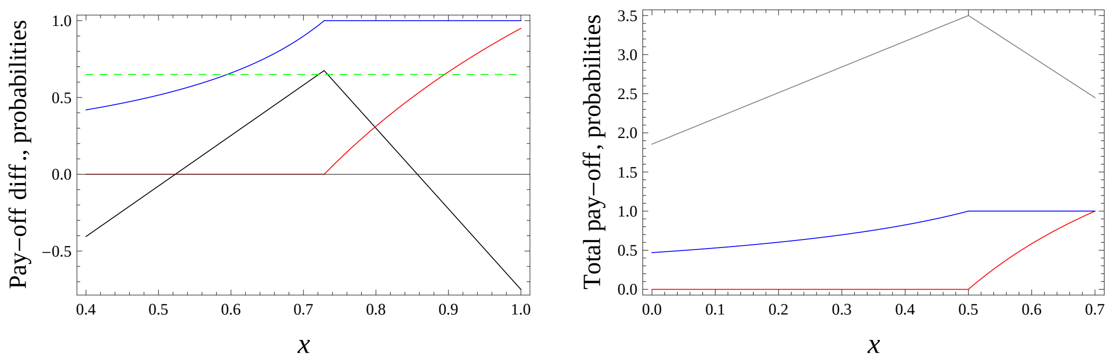
$$\mathfrak{L} = xT(1-P) + P - \Delta(T-1) + T(T-P-1)(1-x)p_2^*, \tag{24}$$
$$p_2^* = \vartheta[\,T-1-P\,]\min\left[\frac{\Delta + xP}{(1-x)(T-P)},\, 1\right], \tag{25}$$
$$\mathfrak{F} = \Delta + P, \tag{26}$$

where $\vartheta[x]$ is the step function: $\vartheta[x<0] = 0$ and $\vartheta[x>0] = 1$. Eq. (26) for the average pay-off $\mathfrak{F}$ of F is obtained directly from (22), where p_2^* is the optimal value of p_2. The corresponding optimal value of p_1^* is found from (22). Thus Δ is a parameter that governs the difference between the actualy pay-off of F and its guaranteed value; see (26). For $x = 1$, we have $\Delta = \delta$ and (22, 24, 26) revert to (resp.) (4, 5, 6).

One way of dealing with (24, 25) is to maximize $\mathfrak{L}(x)$ over x keeping Δ fixed, i.e. keeping the average pay-off of F fixed. To this end, we look at $\mathfrak{L}(x)$ for $x \lesssim 1$. This shows that for $T < 2 = 2R$, function $\mathfrak{L}(x)$ locally maximizes at $x = 1$, which is clearly also the global maximum of $\mathfrak{L}(x)$. Putting $x = 1$ into (21–26) we revert to the previous situation; see (4–6). It is clear from the structure of (20, 21) that after maximization over x we produce a Pareto-optimal solution: $\mathfrak{L}_c$ and $\mathfrak{F}_c$ in (5, 6) are such that neither can be increased by not decreasing another; see Fig. 1 for illustration [6].

---

[6] More formally, we get a Pareto-optimal solution by keeping one (average) pay-off fixed and then maximizing over another [66]. This is because both pay-offs are concave (moreover linear) varying over a convex set [66].



(a) (b)

FIG. 2: These figures illustrate (24–26) for the prisoner's dilemma with $T > 2R$ and compare them with the situation $T < 2R$; see (1, 2).
(a) The pay-off difference $\mathfrak{L} - \mathfrak{L}_c$ (black curve) given by (26, 5) versus x for the prisoner's dilemma game (1, 2) with $T = 3.5$, $R = 1$, $P = 0.2$, $\Delta = 0.75$ and $\delta = 0.45$; cf. (14). Red and blue curves: the probabilities p_1^* and p_2^* (resp.) given by (14, 22 25). Dashed green line: $P + \delta$; cf. (4). With these parameters we hold $\mathfrak{L}_c > T/2$; cf. (9).
(b) Gray curve: the summary pay-off $\mathfrak{L} + \mathfrak{F}$ given by (24, 26) for same parameters as in (a), but $\Delta = \frac{T}{2} - P$. The maximal value of $\mathfrak{L} + \mathfrak{F}$ equals T. For this specific choice of Δ, L and F get equal pay-off: $\mathfrak{L} = \mathfrak{F} = T/2$; see (24, 26).

C. The case $T > 2R$

Let us now apply (22–26) to the case where

$$T > 2 = 2R, \tag{27}$$

in (1). This situation is frequently omitted from the consideration of the prisoner's dilemma game, but there are straightforward examples, where it can be realized [7]. We note that under (27), function $\mathfrak{L}(x)$ locally minimizes at $x = 1$, and maximizes at $p_2^* = 1$, i.e. whenever two arguments of max[..., ...] in (25) are equal; see Figs. 2 for illustration. This leads to the value of $x = x_{\max}$ that maximizes $\mathfrak{L}$ for a fixed Δ:

$$x_{\max} = 1 - (P + \Delta)/T, \tag{28}$$

and which automatically holds (23). Then we get from (24, 26, 28):

$$\mathfrak{L}_{\max} \equiv \mathfrak{L}(x_{\max}) = T - P - \Delta = T - \mathfrak{F}, \tag{29}$$

$$p_1^*(x_{\max}) = 0, \qquad p_2^*(x_{\max}) = 1. \tag{30}$$

Note from (29) that the summary (average) pay-off is now equal to its maximal value. In this context, recall from the prisoner's dilemma game (1) under condition (27) that the deterministic one-shot game has the maximal summary pay-off $2R = 2$, but the average maximal summary pay-off is $T > 2R = 2$. Next, (30) shows that the solution is in a sense fair, since L responds by c to d, and by d to c; cf. (4). Indeed, within (29) the pay-off of F need not be smaller than that of L. In particular, we gets equal pay-offs $\mathfrak{L} = \mathfrak{F} = T/2$ for $x = 1/2$. More generally, besides holding (23) and $x_{\max} < 1$, Δ should be such that $\mathfrak{L}_{\max} > P$, which we shall assume to be the case. As seen in Fig. 3, $\mathfrak{L}_{\max}$ and $\mathfrak{F}$ fill the Pareto line on the pay-off diagram in the regime $T > 2 = 2R$.

We now discuss how the transition from $x = 1$ to $x < 1$ can take place. The difference between pay-offs reads from (29, 26, 5, 6):

$$\mathfrak{L}_{\max} - \mathfrak{L}_c = P(T - 2) + \delta(T - 1 - \frac{\Delta}{\delta}), \tag{31}$$

$$\mathfrak{F} - \mathfrak{F}_c = \Delta - \delta. \tag{32}$$

---

[7] Consider two firms (L and F) that sell similar goods. Actions c (d) refer to advertising (not advertising) the goods. If both do not advertise, then they share the market and sell equal amounts. If both advertise, then they mutually neutralize each other, again share the market, but now waste resources for advertising. However, if one advertises and another does not, the one that advertised can take the lion's share of the market. Now condition (27) can be realized in that situation. Moreover, (30) is a type of cartel agreement, where L does advertise, if F did not, and *vice versa*.

It is seen that for $T > 2$ there are situations, for a sufficiently small $\frac{\Delta}{\delta} > 1$, where both $\mathfrak{L}_{\max} > \mathfrak{L}_c$ and $\mathfrak{F} > \mathfrak{F}_c$ can hold, i.e. both L and F can benefit out of allowing F to defect. For $T < 2$ such situations are excluded, as seen from (31). Figs. 2 demonstrate the behavior of $\mathfrak{L}(x)$ for pertinent values of x.

We return to (4–6) and note that they can dissatisfy F due to two reasons: (a) F is not supposed to defect due to $\Pr[\mathrm{L} \mapsto d \,|\, \mathrm{F} \mapsto d] = 1$; cf. (4). (b) His pay-off is small: $\mathfrak{L}_c > \mathfrak{F}_c$. In response, L can propose the set-up (14) that—though being more complex—is beneficial for both for the assumed $T > 2$ case. Fig. 2(a) shows that there is a range of $x \in (x_1, x_2)$, where both $\mathfrak{L} > \mathfrak{L}_c$ and $\mathfrak{F} > \mathfrak{F}_c$ hold. (Obviously, $\mathfrak{L} - \mathfrak{L}_c > 0$ cannot hold for $x = 0$ or $x = 1$.) For $x \in (x_1, x_2)$, p_2 is well above zero, i.e. defections of F are tolerated. Hence $x \in (x_1, x_2)$ is a regime that F can accept: he is not punished for defection, and his pay-off increased compared to $x = 1$. L can offer this regime to F, since his pay-off also increases: $\mathfrak{L} > \mathfrak{L}_c$.

Note that a conscientious F may wish to accept $\Pr[\mathrm{L} \mapsto d \,|\, \mathrm{F} \mapsto d] = 1$ as something rightful, but would like to increase $p_1 = \Pr[\mathrm{L} \mapsto c \,|\, \mathrm{F} \mapsto c]$ from the value $P + \delta$ (at $x = 1$); see (4–6). Now using (22, 24, 25) we obtain

$$\begin{aligned} \mathfrak{L} - \mathfrak{L}_c &= -(T-1)(\Delta - \delta) + T(1 - x[(T - P - 1)p_2^* + P - 1]) \\ &= -(T-1)x(\,p_1^* - [P + \delta]\,) + (1-x)\,[\,(T-1)\delta + T(P-1) - p_2^* P\,]\,, \end{aligned} \tag{33}$$

where $(T-1)\delta + T(P-1) < 0$ due to (4). Hence (33) show that $p_1 > P + \delta$ always (i.e. both for $T < 2R$ and $T > 2R$) leads to $\mathfrak{L} < \mathfrak{L}_c$. An example of this is seen in Fig. 2(a), where $\mathfrak{L}(x) < \mathfrak{L}_c$ whenever $p_1 > P + \delta$.

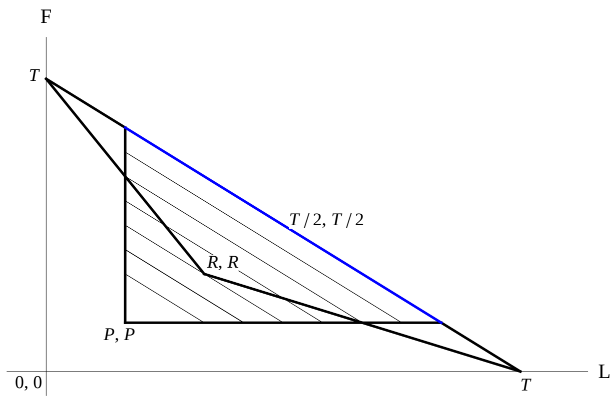


FIG. 3: The same as in Fig. 1, but with $T = 3, R = 1, P = \frac{1}{2}$, i.e. in the regime (27). All pay-offs on the blue line are achieved under certain values of Δ and x related via (28).

V. DISCUSSION: OPEN PROBLEMS OF LEADERSHIP

Our results provide model-dependent answers to several open problems of the leadership research. We reproduce these questions from literature keeping their original form whenever possible.

1: *Not do leaders make a difference, but under what conditions does leadership matter [10]?*

The above leadership scenario provides a solution to the prisoner's dilemma, i.e. the leadership in the model has a functional role. The leader (L) here does have a power over the follower (F), since he convinced F not play the dominant strategy. Recall that the power of L over F is defined by the extent L can get F to do things that F would otherwise not do [53]. The autocratic (coercive or Stackelberg's) leader is not powerful in the (single-shot) prisoner's dilemma game, because he will never change the behavior of the follower who has a dominant strategy, i.e. a strategy that is the best-response to all strategies of the leader.

2: *Are good and poor leadership qualitatively different phenomena [10]?*

The model shows a *qualitative* difference between good (=egalitarian) vs. poor (=exploiting) leadership: the summary pay-off—that determines competitive abilities of the group as a whole—is higher or equal for the egalitarian leadership.

3: *Under what conditions leaders exploit followers [5–8]?*

Within this model, exploitation—where the dilemma is resolved, but L gets more than F—has certain rational roots, because the exploitation regime has a larger stability domain in the sense explained in sections II B and II C. Within the repeated-game implementation of the scenario, the stability relates to back-reaction of F on L, and this does imply costs for F. Another stability mechanism (within the one-shot implementation) relates to the prestige (reputation) of F.

4: *What are game-theoretic formalizations of the Machiavellianism [43]?*

The behavior of L does resemble the Machiavellian personality extensively discussed in literature [43, 44]. This personality is defined by three features: the ability to manipulate and exploit people without provoking them. The Machiavellian personality is considered as a possible model for social intellect [43, 44]. First of all, we note that almost all distinguishing features of manipulation [40–42]—where L and F are (resp.) the manipulator and manipulee—are present with the above leadership scenario:

(1.) The manipulee does keep the full freedom of will. Indeed, in our model F does have the right of the first move. There is no coercion on F.

(2.) F need not be directly deceived [41]. In our model F can know beforehand about all the pay-offs and the rules of the game, which will be followed strictly by L.

(3.) The manipulator lets the manipulee to succumb to a weakness [40]. In our model L can legitimize the exploitation by arguing that it is more stable than the egalitarian regime; see the discussion in sections II B and II C. F may agree with this being afraid to loose his above-guaranteed pay-off (the weakness).

One of defining features of the Machiavellian personality is also present in the model:

(4.) Not provoking (skillfulness): L is ready to accept some limited control from F; see section II C. In a different scenario (realized for $T > 2R$): if F protests against defections of L, the latter can re-manipulate the situation and permit F to defect with a well-chosen probability thereby increasing his own pay-off; see section IV.

Hence the exploiting leadership can be a model for Machiavellian personality, which so far was not properly formalized within game theory. Ref. [43] directly related this personality with the defection strategy of the prisoner's dilemma game, but this relation does not account for its manipulative aspects and hence does not explain why it is a form of social intellect. The defection strategy of the prisoner's dilemma game is easily neutralized by defecting in response.

5: *What are implications of the Machiavellian leadership for followers [8]?*

Since F is subject to manipulation and exploitation, he has to be non-Machiavellian. Such people score low in corresponding tests [43, 44]. Note that the back-reaction of F on L that involves sequential defections of L, also roughly coincides with the behavior of non-Machiavellian followers.

We studied two implications of the exploitative leadership for F. First, F can fall into apathy, because he is put into a situation, where either action by him gives comparable average pay-offs. Another option is that F will be actively dissatisfied by rules of the game, i.e. by the fact that he gets less than L and he is effectively prevented from defecting, while L does defect. Now L can solve both these problems simultaneously—and yet to increase his own average pay-off—by allowing F to defect with a certain (well-chosen) probability. This demands that the pay-off for defecting a cooperating opponent is sufficiently large; see section IV.

There is an evidence that dyadic groups composed of one Machiavellian and one non-Machiavellian are outperformed by two Machiavellians, or two non-Machiavellians that are able to develop more egalitarian coalitions [43]; cf. **2**.

6: *Not what are the traits of leaders, but how do leaders' personal attributes interact with situational properties to shape outcomes [10]?*

In the studied game-theoretic model, L does have personal attributes. For $T < 2R$, the leader L has a manipulative (Machiavellian [43, 44]) social intellect that allows him to gain an additional pay-off via probabilistic response, but L is also willing to admit a limited control from F (via the repeated-game implementation), or from his environment via prestige (reputation) factors. The final pay-off of L is determined by the inter-play between these mechanisms. Now these attributes are optimal (i.e. emerge out of pay-off maximization) for the prisoner's dilemma game considered. For other symmetric games they are inferior: the prisoner's dilemma (1) is the only symmetric game that admits a non-trivial probabilistic response (4) [54]. E.g. for symmetric coordination games a probabilistic response can be introduced, but it is inferior to purely deterministic responses [54].

For $T > 2R$ (see section IV), the leader L should deal with the complexity of the situation, since L should to some extent understand motivations of the follower F (i.e. why F is defecting). Then it is possible to improve both pay-offs of L and F.

7: *What can be said about the situation with many followers and more than one leader [16]?*

This interesting problem should be studied in future possibly via tools of statistical mechanics that were recently applied to game theory [67–69].

Acknowledgment

We thank A. Khachatryan for useful discussions. AEA and SGB were supported by SCS of Armenia, grants No. 18RF-002 and No. 18T-1C090. AVM was supported by RFBR via the research project 18-51-05007 arm_a.

Appendix A: Extended representation of the probabilistic commitment

Let us write a generic 2×2 game as

L/F	α	β
i	$s_{i\alpha},\, \sigma_{i\alpha}$	$s_{i\beta},\, \sigma_{i\beta}$
j	$s_{j\alpha},\, \sigma_{j\alpha}$	$s_{j\beta},\, \sigma_{j\beta}$

, (A1)

where notations obviously generalize (1).

Let now the game (A1) be considered in the sequential form: F makes the first move with the strategies α or β, while the strategies of L are announced beforehands, i.e. before knowing the move of F. Hence there are four of them: ii and jj mean that L will act (resp.) i and j for any move of F, ij means that L will act i (j) in response to α (β), while ji means that L will act j (i) in response to α (β). We represent the situation as follows:

L/F	α	β
ii	$s_{i\alpha},\, \sigma_{i\alpha}$	$s_{i\beta},\, \sigma_{i\beta}$
ij	$s_{i\alpha},\, \sigma_{i\alpha}$	$s_{j\beta},\, \sigma_{j\beta}$
ji	$s_{j\alpha},\, \sigma_{j\alpha}$	$s_{i\beta},\, \sigma_{i\beta}$
jj	$s_{j\alpha},\, \sigma_{j\alpha}$	$s_{j\beta},\, \sigma_{j\beta}$

(A2)

Let us now endow L and F with mixed strategies: the four strategies of L get probabilities P_{ii}, P_{ij}, P_{ji}, and P_{jj} with $P_{ii} + P_{ij} + P_{ji} + P_{jj} = 1$. Note that out of these quantities we can construct conditional probabilities:

$$P_{ii} + P_{ij} = P_{i|\alpha}, \quad P_{jj} + P_{ji} = P_{j|\alpha}, \quad P_{ii} + P_{ji} = P_{i|\beta}, \quad P_{jj} + P_{ij} = P_{j|\beta}. \tag{A3}$$

The two strategies of F have probabilities π_α and π_β with $\pi_\alpha + \pi_\beta = 1$.

Employing the prisoner's dilemma game pay-offs (1) in (A2), we get (d, dd) is the only Nash equilibrium of the game. Putting into (A1) pay-offs of the ultimatum game as defined in section III B we obtain:

II/I	f	u
a	$(0.5\,\mathfrak{M},\, 0.5\,\mathfrak{M})$	$(0.01\,\mathfrak{M},\, 0.99\,\mathfrak{M})$
r	$(0\,,0)$	$(0\,,0)$

, (A4)

where a and r (f and u) are actions of II (I). Recall that II makes the second move. There is a single Nash equilibrium in (A4): (u, a). Now using (A4) in (A2), we shall get two equilibria: (u, aa) (that resembles the previous one), and (f, ar), where II treatens I to reject the unfair offer. This second Nash equilibrium is not sub-game perfect (credible), since r is never the best response of II.

Appendix B: Second-order metagames

Here we shall present the second-order metagame approach [20, 37] in a slightly generalized, probabilistic form that will make obvious its main assumption. Given the game (A1), let us define conditional probabilities $P^{[a]}_{i|\alpha} = 1 - P^{[a]}_{j|\alpha}$ and $P^{[a]}_{i|\beta} = 1 - P^{[a]}_{j|\beta}$ for the response of L to pure strategies of F. The index $a = 1, 2, .., K_{\rm L}$ refers to different possible responses (i.e. meta-strategies of L), their number and form need not be specified for our present purposes. Then the first-order metagame is defined by average pay-offs:

$$\widetilde{s}_{a\alpha} \equiv P^{[a]}_{i|\alpha} s_{i\alpha} + P^{[a]}_{j|\alpha} s_{j\alpha}, \qquad \widetilde{\sigma}_{a\alpha} \equiv P^{[a]}_{i|\alpha} \sigma_{i\alpha} + P^{[a]}_{j|\alpha} \sigma_{j\alpha}, \tag{B1}$$

for L and F, respectively. The meaning of (B1) is clear and refers to averages over pure strategies of L, given the probabilistic response of L to pure strategies of F.

Starting from (B1), Howard [37] (see [20] for a review) proceeds to define the second-order meta-game approach, where F now probabilistically reacts to meta-strategies of L. These reactions are described by conditional probabilities $P^{[b]}_{\alpha|a}$, where $b = 1, 2, .., K_{\rm F}$. Although Howard restricted the situation to pure (non-mixed) responses $P^{[b]}_{\alpha|a} = 0, 1$, this restriction is not essential for the argument we want to make.

This invites to define the second-order pay-offs by analogy to (B1):

$$\widetilde{\widetilde{s}}_{ab} \equiv \widetilde{s}_{a\alpha} P^{[b]}_{\alpha|a} + \widetilde{s}_{a\beta} P^{[b]}_{\beta|a}, \qquad \widetilde{\widetilde{\sigma}}_{ab} \equiv \widetilde{\sigma}_{a\alpha} P^{[b]}_{\alpha|a} + \widetilde{\sigma}_{a\beta} P^{[b]}_{\beta|a}. \tag{B2}$$

Separate terms in (B2) are interpreted as follows; e.g. $s_{i\alpha} P^{[a]}_{i|\alpha} P^{[b]}_{\alpha|a}$ that enters into the first equation in (B2) means that L first announces his reaction a (commits to a), then F acts in response α with probability $P^{[b]}_{\alpha|a}$, and then L acts i with the promised reaction (probability) $P^{[a]}_{i|\alpha}$. Thus the strategies of L (F) amount to $a = 1, 2, .., K_{\rm L}$ ($b = 1, 2, .., K_{\rm F}$). It is important to stress here that the consistency of the last move of L with its own commitment is postulated here. The credibility (i.e. stability) of such a commitment was studied in [59].

[1] *Encyclopedia of leadership*, G.R. Goethals, G.J. Sorenson, J.M. Burns, editors (Thousand Oaks, California: Sage; 2004).
[2] R. L. Ackoff, *Transformational leadership*, Strategy & Leadership, **27**, 20-25 (1999).
[3] J. Antonakis, D. V. Day and B. Schyns, *Leadership and individual differences: At the cusp of a renaissance*, The Leadership Quarterly, **23**, 643-650 (2012).
[4] P.D. Harms, D. Wooda, K. Landay, P. B. Lester, and G. V. Lester, *Autocratic leaders and authoritarian followers revisited: A review and agenda for the future*, The Leadership Quarterly **29**, 105-122 (2018).
[5] A. J. King, D. D.P. Johnson, and M. Van Vugt, *The Origins and Evolution of Leadership*, Current Biology 19, R911R916 (2009).
[6] P. L. Hooper, H. S. Kaplan, and J. L. Boone, *A theory of leadership in human cooperative groups*, Journal of Theoretical Biology, **265**, 633-646 (2010).
[7] C. Koykka and G. Wild, *The evolution of group dispersal with leaders and followers*, Journal of Theoretical Biology, **371**, 117-126 (2015).
[8] J.E. Smith *al.*, *Leadership in Mammalian Societies: Emergence, Distribution, Power, and pay-off*, Trends Ecol. Evol. **31**, 54-66, (2016).
[9] R. Hogan and R. B. Kaiser, *What We Know About Leadership*, Review of General Psychology, **9**, 169180 (2005).
[10] J.R. Hackman and R. Wageman, *Asking the right questions about leadership*, American Psychologist, **62**, 43-47 (2007).
[11] V. M. Eguiluz, M. G. Zimmermann, C. J. Cela-Conde, and M. San Miguel, *Cooperation and the Emergence of Role Differentiation in the Dynamics of Social Networks*, American Journal of Sociology, **110**, 977-1008 (2005).
[12] M.G. Zimmermann and V.M. Eguiluz, *Cooperation, social networks, and the emergence of leadership in a prisoner's dilemma with adaptive local interactions*, Physical Review E, **72**, 056118 (2005).
[13] J.K. Hazy, *Computer models of leadership: Foundations for a new discipline or meaningless diversion?* The Leadership Quarterly, **18**, 391-410 (2007).
[14] D. Pais and N.E. Leonard, *Adaptive network dynamics and evolution of leadership in collective migration*, Physica D, **267**, 81-93 (2014).
[15] S. Galam, *Sociophysics: A Physicist's Modeling of Psycho-political Phenomena* (Springer, Berlin, 2012).
[16] A. E. Allahverdyan and A. Galstyan, *Emergence of Leadership in Communication*, PLoS ONE **11**, e0159301 (2016).
[17] R. L. Calvert, *Leadership and Its Basis in Problems of Social Coordination*, International Political Science Review, **13**, 7-24 (1992).
[18] J.M. Colomer, Rationality and Society, **7**, 225 (1995).
[19] R.D. Luce and H. Raiffa, *Games and decisions: Introduction and critical survey* (Wiley, Oxford, England, 1957).
[20] M. Shubik, *Game Theory, Behavior and the Paradox of the Prisoner's Dilemma: Three Solutions*, Journal of Conflict Resolution, **14**, 181 (1970).
[21] R.B. Myerson, *Game Theory: Analysis of Conflict* (Harvard University Press, MA, 1997).
[22] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, 1998).
[23] *The Prisoner's Dilemma*, M. Peterson (ed.) (Cambridge University Press, Cambridge, 2015).
[24] E. Sober and D. S. Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Harvard University Press, Cambridge, MA, 1998).
[25] D. G. Arce M. and T. Sandler, *The Dilemma of the Prisoners' Dilemmas*, KYKLOS, **58**, 324 (2005).
[26] W. H. Press and F. J. Dyson, *Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent*, PNAS, **109**, 1040910413 (2012).
[27] J. Liu *et al.*, *Evolutionary behavior of generalized zero-determinant strategies in iterated prisoner's dilemma*, Physica A **430**, 81-92 (2015).
[28] W. T. Bianco and R. H. Bates, *Cooperation by Design: Leadership, Structure, and Collective Dilemmas*, The American Political Science Review, **84**, 133-147 (1990).
[29] E. Akin, *The Iterated Prisoner's Dilemma: Good Strategies and Their Dynamics*, arXiv:1211.0969.
[30] H. von Stackelberg, *The Theory of Market Economy* (Oxford University Press, Oxford, 1952).
[31] B. Von Stengel and S. Zamir, *Leadership with commitment to mixed strategies*, research report LSE-CDAM-2004-01, London School of Economics, 2004.
[32] V. Conitzer, *On Stackelberg Mixed Strategies*, Synthese, **193**, 689-703 (2016).

[33] Z. Wang, L. Wang, A. Szolnoki, M. Perc, *Evolutionary games on multilayer networks: a colloquium*, Eur. Phys. J. B **88**, 124 (2015).
[34] M. Perc, A. Szolnoki, *Coevolutionary games: a mini review*, BioSystems **99**, 109-125 (2010).
[35] M. Perc, J.J. Jordan, D. G. Rand, Z. Wang, S. Boccaletti, and A. Szolnoki, *Statistical physics of human cooperation*, Physics Reports, **687**, 1-51 (2017).
[36] G. Ver Steeg and A. Galstyan, *Information transfer in social media*, in Proceedings of the 21st international conference on World Wide Web, 509-518 (2012). See also arXiv:1110.2724.
[37] N. Howard, *Paradoxes of Rationality: Theory of Metagames and Political Behaviour* (MIT Press, 1971).
[38] G.J. Olsder, *Phenomena in Inverse Stackelberg Games, Part 1: Static Problems*, Journal of Optimization Theory and Applications, **143**, 589 (2009).
[39] N. Groot, B. De Schutter, and H. Hellendoorn, *Reverse stackelberg games, part I: Basic framework*, in IEEE International Conference on Control Applications (CCA), pp. 421-426 (2012).
[40] A. Barnhill, *What Is Manipulation?*, in *Manipulation: Theory and Practice* edited by C. Coons and M. Weber (Oxford Scholarship Online, 2014); DOI:10.1093/acprof:oso/9780199338207.003.0003.
[41] J. Rudinow, *Manipulation*, Ethics **88**, 338-347 (1978).
[42] E.L. Dotsenko, *Psychology of manipulation: phenomena, mechanisms and protection* (Moscow, CheRo, 1997).
[43] D. S. Wilson, D. Near, and R. R. Miller, *Machiavellianism: A Synthesis of the Evolutionary and Psychological Literatures*, Psychological Bulletin, **119**, 285-299 (1996).
[44] T. Bereczkei, *Machiavellianism: The Psychology of Manipulation* (Routledge, NY, 2018).
[45] T.C. Schelling, *The Strategy of Conflict* (Harvard University Press, Cambridge, 1960).
[46] *Evolution and the capacity for commitment*, ed. by R.M. Nesse (Russell Sage Press, NY, 2001).
[47] S.J. Brams, *Newcomb's Problem and Prisoner's Dilemma*, Journal of Conflict Resolution, **19**, 596 (1975).
[48] S.J. Brams and M. Kilgour, *Stabilizing unstable outcomes in prediction games*, MPRA Paper No. 77655 (2017).
[49] E. Fehr, *Dont lose your reputation*, Nature, **432**, 449450 (2004).
[50] M. Milinski, *Gossip and reputation in social dilemmas*, in The Oxford Handbook of Gossip and Reputation, p.193, edited by F. Giardini and R.P.M. Wittek (Oxford University Press, Oxford 2018).
[51] C. Roddie, *Reputation and gossip in game theory*, in The Oxford Handbook of Gossip and Reputation, p.193, edited by F. Giardini and R.P.M. Wittek (Oxford University Press, Oxford 2018).
[52] Han-Xin Yang and Jing Yang, *Reputation-based investment strategy promotes cooperation in public goods games*, Physica A (2019) in press, https://doi.org/10.1016/j.physa.2019.04.190
[53] W.H. Riker, *Some ambiguities in the notion of power*, American Political Science Review, **58**, 341 (1964).
[54] S. G. Babajanyan and A. E. Allahverdyan, in preparation.
[55] M. Grigoryan, *Armenia's anti-corruption council accused of lavish spending*, The Guardian, 12 August 2015.
[56] J. Macrae, *Underdevelopment and the Economics of Corruption: A Game Theory Approach*, World Development **10**, 677 (1982).
[57] C. Van Rijckeghem and B. Weder, *Corruption and the Rate of Temptation: Do Low Wages in the Civil Service Cause Corruption?*, International Monetary Fund Working Paper, WP/97/73 (1997).
[58] See in https://transparency.am/en/cpi
[59] L. Renou, *Commitment games*, Games and Economic Behavior **66**, 488-505 (2009).
[60] R.J. Aumann and L. S. Shapley. *Long-term competition: a game-theoretic analysis*, in *Essays in Game Theory in Honor of Michael Maschler*, ed. by N.Megiddo (Springer-Verlag, New York, 1994) pp. 115.
[61] D. Fudenberg and E. Maskin, *The folk theorem in repeated games with discounting or with incomplete information*, in *A Long-Run Collaboration On Long-Run Games* ed. by D. Fudenberg and D. K. Levine (World Scientific Publishing Co., 2008), pp 209-230.
[62] J. Gale, K.G. Binmore, and L. Samuelson, *Learning to be Imperfect: The Ultimatum Game*, Games and Economic Behavior, **8**, 5690 (1995).
[63] B. Skyrms, *Evolution of the Social Contract* (Cambridge University Press, Cambridge, 1996).
[64] H. T Anh, F. C. Santos, T. Lenaerts, and L. M. Pereira, *Synergy between intention recognition and commitments in cooperation dilemmas*, Sci. Rep. **5** 9312 (2015).
[65] Y. Fujimoto and K. Kaneko, *Functional dynamic by intention recognition in iterated games*, New J. Phys. **21**, 023025 (2019).
[66] S. Karlin, *Mathematical Methods and Theory in Games, Programming, and Economics* (Pergamon Press, London, 1959).
[67] G. Szabo and I. Borsos, *Evolutionary potential games on lattices*, Physics Reports, **624**, 1 (2016).
C. Hauert and G. Szabo, *Game theory and physics*, American Journal of Physics, **73**, 405-414 (2005).
[68] C. Adami and A. Hintze, *Thermodynamics of Evolutionary Games*, Physical Review E **97**, 062136 (2018).
[69] C. Benjamin and S. Sarkar, *Triggers for cooperative behavior in the thermodynamic limit: A case study in Public goods game featured*, Chaos **29**, 053131 (2019).